VIROLOGY 157, 31–39 (1987)

Hantaan Virus M RNA: Coding Strategy, Nucleotide Sequence, and Gene Order

CONNIE S. SCHMALJOHN,*,[1] ALAN L. SCHMALJOHN,† AND JOEL M. DALRYMPLE*

*Virology Division, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland 21701-5011, and †Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland 21201

Received August 25, 1986; accepted October 21, 1986

The M genome segment of Hantaan virus was molecularly cloned and the nucleotide sequence of cDNA was determined. The virion RNA is 3616 bases long with 3'- and 5'-terminal nucleotide sequences complementary for 18 bases. A single long open reading frame in the viral complementary-sense RNA had the potential to encode 1135 amino acids or a polypeptide of 126,000 Da. Amino-terminal sequences of isolated G1 and G2 envelope glycoproteins were determined, revealing a gene order with respect to message sense RNA of 5'-G1-G2-3'. Mature G1 begins 18 amino acids beyond the first AUG of the open reading frame, preceded by a short, hydrophobic leader sequence. G2 begins at the 649th amino acid of the open reading frame and also follows a hydrophobic sequence. Carboxy termini of G1 and G2 were localized and gene order was verified by immune precipitation of Hantaan proteins with antisera to synthetic peptides generated by using amino acid sequences derived from the cDNA sequence. The antipeptide sera were also reactive by immunoblotting with SDS-denatured G1 and G2. Molecular weights of 64,000 and 53,700 were calculated for the G1 and G2 glycoproteins, respectively, from their predicted amino acid sequences. Five potential asparagine-linked glycosylation sites were contained within the G1 amino acid sequence and two within the G2 sequence. These data are consistent with our previous estimates of the molecular weights and extent of glycosylation of the Hantaan envelope glycoproteins.

INTRODUCTION

Hantaan virus is the type species of the *Hantavirus* genus of the Bunyaviridae family. Hantaviruses have been associated with many clinically similar diseases known collectively as hemorrhagic fever with renal syndrome. Hantaan virus is the etiologic agent of Korean hemorrhagic fever, which is one of the better known and most severe diseases included in the syndrome. Like other Bunyaviridae, Hantaan possesses a three-segmented, single-stranded RNA genome of negative polarity (Schmaljohn and Dalrymple, 1983). The large (L), medium (M), and small (S) genome segments have relative molecular masses (M_r) of approximately 2.7, 1.2, and 0.6 × 10^6, respectively, as estimated by agarose gel electrophoresis, and are enclosed in three separate nucleocapsid structures which are surrounded by a lipid envelope containing two virus-specified glycoproteins designated G1 and G2 (Schmaljohn *et al.*, 1983; Schmaljohn and Dalrymple, 1983, 1984; Elliot *et al.*, 1984).

Although the coding strategies of viruses in different genera of the Bunyaviridae have been shown to vary considerably, all viruses examined to date encode their nucleocapsid protein and a nonstructural protein in the S segment and their envelope glycoproteins in the M segment. By deduction, the L segment is believed to encode the virion-associated polymerase (reviewed in Bishop, 1985).

Our studies on the basic molecular mechanisms of Hantaan replication, including the coding strategies of the genome segments, have been directed toward a better understanding of the viral infection process. We have found that Hantaan, like other viruses in the family, encodes its nucleocapsid protein in the S RNA segment (Schmaljohn *et al.*, 1986b). Sequence analysis of cloned cDNA from the S segment revealed only a single long open reading frame (ORF) in the viral complementary sense RNA, which was shown to include the gene for the nucleocapsid protein (Schmaljohn *et al.*, 1986c). The only other potential gene product was a 6K polypeptide (48 amino acids), encoded in the same reading frame as the major ORF, initiating at an AUG three bases beyond the ochre stop codon of the nucleocapsid protein. However, the actual existence of any nonstructural polypeptide has not yet been demonstrated. These data suggested that the coding strategy of the S segment of Hantaan differs considerably from the elaborate ambisense strategy of the phlebovirus S segment and the overlapping reading frame strategy employed by the bunyavirus S segment to encode both their nucleocapsid protein and a nonstructural (NS_s) polypeptide (Bishop, 1985).

Members of the Bunyaviridae for which sequence data of the M genome segment have been reported include two viruses in the *Phlebovirus* genus, Rift Valley

[1] To whom reprint requests should be addressed.

0042-6822/87 $3.00

fever and Punta Toro viruses, and two viruses in the *Bunyavirus* genus, snowshoe hare and Bunyamwera viruses (Collett *et al.*, 1985; Ihara *et al.*, 1985; Eshita and Bishop, 1984; Lees *et al.*, 1986). All of these viruses were found to encode single gene products in the viral complementary-sense RNA. Amino-terminal sequence analysis of the phlebovirus glycoproteins indicated that both G1 and G2 were contained within that gene product, and sufficient coding information preceded the amino terminus of the first encoded glycoprotein to produce a polypeptide of 14K for Rift Valley fever or 30K for Punta Toro viruses. This putative polypeptide, termed NS_M, has not been identified in infected cells, and its significance remains unknown.

Although the gene order has not yet been determined for members of the bunyavirus genus, the M segments of these viruses encode a NS_M polypeptide in addition to G1 and G2. While NS_M proteins have been observed in bunyavirus-infected cells (Fuller and Bishop, 1982), functions for these polypeptides have similarly not been assigned.

To our knowledge, sequence information is not yet available for viruses in the *Uukuvirus* or *Nairovirus* genera. However, the G1 and G2 proteins of Uukuniemi virus have been shown to be processed from a 110K polyprotein precursor in cell-free translation studies (Ulmanen *et al.*, 1981), and G1 was found to be amino terminal and G2 carboxy terminal in the 110K precursor by pulse–chase labeling of infected cells (Kuismanen, 1984).

Therefore, unlike the S genome segments of Bunyaviridae, which demonstrate widely disparate coding strategies among the five genera, the M RNA segments of all characterized viruses in this family appear to encode polyprotein precursors which are processed into envelope glycoproteins. We report here that the M genome segment of Hantaan virus conforms to this general observation. Localization of the proteins within the derived amino acid sequence, however, suggests that Hantaan may differ from other viruses in the processing of its envelope proteins.

MATERIALS AND METHODS

Molecular cloning and sequence analysis of Hantaan virus M RNA

Hantaan virus was propagated in Vero E6 cells (C1008, ATCC), purified by sucrose gradient sedimentation, and RNA was extracted from virions as previously described (Schmaljohn *et al.*, 1983; Schmaljohn and Dalrymple, 1983). First-strand cDNA synthesis was primed by the addition of a synthetic oligonucleotide complementary to 19 bases at the 3′ end of virion M RNA (5′-TAGTAGTAGACTCCGCAAA-3′) (Schmaljohn and Dalrymple, 1983). Double-stranded cDNA was synthesized, tailed, and inserted into the *Pst*I site of pBR322, and recombinant plasmids were identified as previously described (Schmaljohn *et al.*, 1986c). Sequence analysis was performed by the dideoxy chain termination method of Sanger *et al.*, (1977), using synthetic oligonucleotides to prime synthesis on cDNA subcloned into the M13 bacteriophage (Messing, 1983). Sequence data were analyzed with the Intelligenetics sequence analysis program on a DEC VAX 11/750 computer or with the University of Minnesota sequence analysis program and an Apple IIe computer.

Amino-terminal sequence analysis of G1 and G2

Hantaan G1 and G2 glycoproteins were prepared for amino-terminal sequence analysis by electrophoresis through polyacrylamide gels by the method of Dietzschold *et al.* (1986). Proteins were boiled in 65 m*M* Tris–borate, pH 8.4, containing 1% SDS, 5% 2-mercaptoethanol, and 10% glycerol and applied to a 3-mm-thick, discontinuous, polyacrylamide gel. A 60-ml separating gel of 10% acrylamide and 0.05% bisacrylamide was prepared in 0.375 *M* Tris–sulfate, pH 8.4. A 20-ml stacking gel of 7% acrylamide and 0.35% bisacrylamide was prepared in 0.17 *M* Tris–sulfate, pH 8.4. Gels were allowed to polymerize for at least 24 hr and were electrophoresed at 100 V for approximately 12 hr, after which they were sliced horizontally. Proteins were recovered from gel slices by agitation in 50 m*M* ammonium bicarbonate containing 0.1% SDS for two separate 12-hr periods. Samples were lyophilized and dialyzed extensively against water. Approximately 100 pmol of purified proteins was sequenced on Model 470A protein sequencer (Applied Biosystems Inc., San Francisco, CA) and analyzed with a 120A on-line PTH analyzer.

Peptide synthesis, immune precipitation, and immunoblot analysis

Peptides representing derived amino acid sequences were synthesized with an Applied Biosystems, Inc. Model 430A peptide synthesizer. Peptides were cleaved from the resin with hydrofluoric acid (Tam and Merrifield, 1985; Stewart and Young, 1984) and coupled to keyhole limpet hemocyanin (KLH) (Liu *et al.*, 1979).

Approximately 1–1.5 mg of each coupled peptide was injected intradermally with Freund's complete adjuvant into four shaved sites on each of two young

female New Zealand white rabbits. Rabbits were boosted with approximately 1.5 mg of coupled peptides in Freund's incomplete adjuvant at 2-week intervals and bled from 10 to 14 days postboost. Rabbit sera were preadsorbed with uninfected Vero E6 cells prior to use in Hantaan immune-precipitation experiments. Polyclonal rabbit antisera consisted of convalescent sera collected 4–12 weeks following infection of New Zealand White rabbits with Hantaan virus (strain 76-118). Intracellular viral proteins were radiolabeled with [^{35}S]methionine (100 μCi/25-cm^2 flask of confluent Vero E6 cells) from 24 to 48 hr postinfection. Cells were lysed on ice in buffer containing 10 m*M* Tris–HCl, pH 8.0; 1 m*M* EDTA; 0.5*M* NaCl; 0.25 mg/ml each of aprotinin and α_2-macroglobulin (Sigma, St. Louis, MO); and 4% Zwittergent 3-14, (Calbiochem-Behring, San Diego, CA). Cell nuclei were removed by centrifugation at 13,000 *g* for 5 min at 4°. Virus-specific proteins were immune precipitated from infected cell lysates by incubation with antisera on ice overnight followed by the addition of 100 μl of 50% protein A–Sepharose (Sigma) and continued incubation for 30 min. Precipitates were washed three times in 0.5× lysis buffer and once in 10 m*M* Tris–HCl, pH 6.7. Electrophoresis was performed at 100 V for 12 hr in SDS-containing polyacrylamide gels as previously described (Schmaljohn *et al.*, 1983).

For immunoblot studies, proteins were electrophoretically transferred from polyacrylamide gels to nitrocellulose membranes (Hoefer Scientific Instruments, San Francisco, CA). Nonspecific sites were blocked by incubation of membranes in a 10% nonfat dry milk solution overnight at room temperature (Johnson *et al.*, 1984) and were probed with antisera (1:50 dilution) for 45 min at 37° in a Decaprobe hybridization chamber (Hoefer Scientific Instruments, San Fancisco, CA). Following a 10-min wash in phosphate-buffered saline (PBS, Sigma), two washes for 20 min each in PBS with 0.05% NP-40 (Bethesda Research Laboratories, Gaithersburg, MD), and one 10-min wash in PBS, 0.5 μCi/lane of ^{35}S-protein A (Amersham, Arlington Heights, IL) was added, and incubation was continued for an additional 30 min at room temperature. Washing was repeated and the membrane was dried and exposed to X-ray film (Kodak XAR, Rochester, NY) at −70°.

RESULTS

Sequence analysis of cDNA and features of Hantaan virus M RNA

Hantaan-specific cDNA inserts were excised from pBR322 plasmids, sized, and restriction mapped. The largest cDNA clone identified appeared of sufficient length to contain most or all of the Hantaan M RNA segment. The nucleotide sequence of this cDNA clone was determined by dideoxy chain termination synthesis on cDNA templates subcloned into bacteriophage M13. Sequences identified, adjacent to the added homopolymeric tail, corresponded to all but the six most 3′ proximal nucleotides reported for Hantaan virion RNA (3′-AUCAUCAUCUGAGGCGUUUUCUUUG) (Schmaljohn *et al.*, 1985). Comparison of the termini of the cDNA clone revealed complementary sequences for 18 bases which included all but the three most distal nucleotides complementary to the virion RNA sequence. Since 3′ and 5′ complementarity is a general feature of the RNA segments of Bunyaviridae (Parker and Hewlett, 1981; Bishop *et al.*, 1982; Cabridilla *et al.*, 1983; Ihara *et al.*, 1984; Collett *et al.*, 1985), we concluded that the large cDNA clone represented the entire M RNA segment except for six bases corresponding to the 3′ terminus and three bases corresponding to the 5′ terminus of virion RNA.

A base-paired stem structure with a calculated free energy of −28.2 kcal/mol was postulated based on this terminal nucleotide sequence complementarity (Fig. 1). A base composition for the entire virion M RNA was determined from cDNA sequence information to be 29.9% A, 17.9% G, 21.4% C, and 30.8% U.

Coding capacity and gene order of the M RNA

A single long open reading frame was detected in the nuclectide sequence corresponding to viral com-

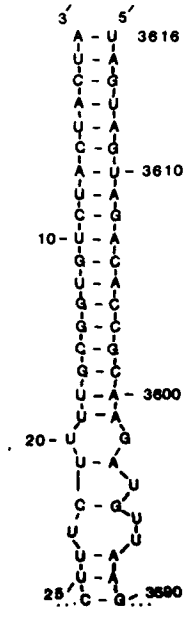
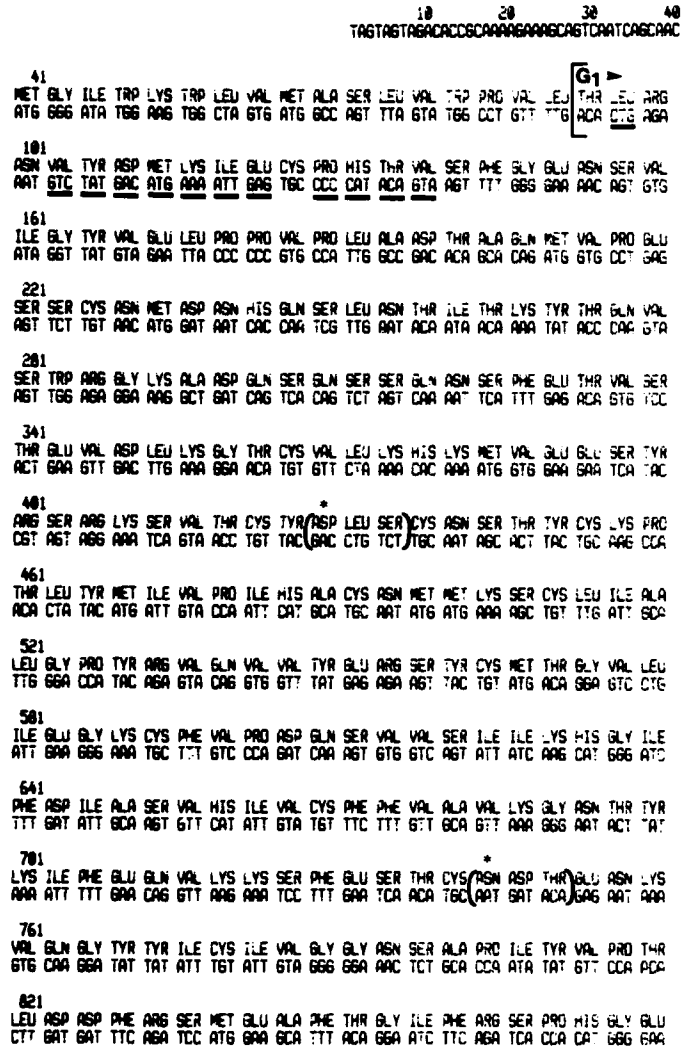


FIG. 1. Predicted secondary structure of the 3′- and 5′-terminal nucleotide sequences of Hantaan M RNA.


10 20 30 40
TAGTAGTAGACACCGCAAAAGAAAGCAGTCAATCAGCAAC
41
MET GLY ILE TRP LYS TRP LEU VAL MET ALA SER LEU VAL TRP PRO VAL LEU THR LEU ARG
ATG GGG ATA TGG AAG TGG CTA GTG ATG GCC AGT TTA GTA TGG CCT GTT TTG ACA CTG AGA
G1
101
ASN VAL TYR ASP MET LYS ILE GLU CYS PRO HIS THR VAL SER PHE GLY GLU ASN SER VAL
AAT GTC TAT GAC ATG AAA ATT GAG TGC CCC CAT ACA GTA AGT TTT GGG GAA AAC AGT GTG
161
ILE GLY TYR VAL GLU LEU PRO PRO VAL PRO LEU ALA ASP THR ALA GLN MET VAL PRO GLU
ATA GGT TAT GTA GAA TTA CCC CCC GTG CCA TTG GCC GAC ACA GCA CAG ATG GTG CCT GAG
221
SER SER CYS ASN MET ASP ASN HIS GLN SER LEU ASN THR ILE THR LYS TYR THR GLN VAL
AGT TCT TGT AAC ATG GAT AAT CAC CAA TCG TTG AAT ACA ATA ACA AAA TAT ACC CAG GTA
281
SER TRP ARG GLY LYS ALA ASP GLN SER GLN SER SER GLN ASN SER PHE GLU THR VAL SER
AGT TGG AGA GGA AAG GCT GAT CAG TCA CAG TCT AGT CAA AAT TCA TTT GAG ACA GTG TCC
341
THR GLU VAL ASP LEU LYS GLY THR CYS VAL LEU LYS HIS LYS MET VAL GLU GLU SER TYR
ACT GAA GTT GAC TTG AAA GGA ACA TGT GTT CTA AAA CAC AAA ATG GTG GAA GAA TCA TAC
401
ARG SER ARG LYS SER VAL THR CYS TYR ASP LEU SER CYS ASN SER THR TYR CYS LYS PRO
CGT AGT AGG AAA TCA GTA ACC TGT TAC GAC CTG TCT TGC AAT AGC ACT TAC TGC AAG CCA
461
THR LEU TYR MET ILE VAL PRO ILE HIS ALA CYS ASN MET MET LYS SER CYS LEU ILE ALA
ACA CTA TAC ATG ATT GTA CCA ATT CAT GCA TGC AAT ATG ATG AAA AGC TGT TTG ATT GCA
521
LEU GLY PRO TYR ARG VAL GLN VAL VAL TYR GLU ARG SER TYR CYS MET THR GLY VAL LEU
TTG GGA CCA TAC AGA GTA CAG GTG GTT TAT GAG AGA AGT TAC TGT ATG ACA GGA GTC CTG
581
ILE GLU GLY LYS CYS PHE VAL PRO ASP GLN SER VAL VAL SER ILE ILE LYS HIS GLY ILE
ATT GAA GGG AAA TGC TTT GTC CCA GAT CAA AGT GTG GTC AGT ATT ATC AAG CAT GGG ATC
641
PHE ASP ILE ALA SER VAL HIS ILE VAL CYS PHE PHE VAL ALA VAL LYS GLY ASN THR TYR
TTT GAT ATT GCA AGT GTT CAT ATT GTA TGT TTC TTT GTT GCA GTT AAA GGG AAT ACT TAT
701
LYS ILE PHE GLU GLN VAL LYS LYS SER PHE GLU SER THR CYS ASN ASP THR GLU ASN LYS
AAA ATT TTT GAA CAG GTT AAG AAA TCC TTT GAA TCA ACA TGC AAT GAT ACA GAG AAT AAA
761
VAL GLN GLY TYR TYR ILE CYS ILE VAL GLY GLY ASN SER ALA PRO ILE TYR VAL PRO THR
GTG CAA GGA TAT TAT ATT TGT ATT GTA GGG GGA AAC TCT GCA CCA ATA TAT GTT CCA ACA
821
LEU ASP ASP PHE ARG SER MET GLU ALA PHE THR GLY ILE PHE ARG SER PRO HIS GLY GLU
CTT GAT GAT TTC AGA TCC ATG GAA GCA TTT ACA GGA ATC TTC AGA TCA CCA CAT GGG GAA
1781
GLU ALA ALA PHE GLN ALA HIS TYR LYS VAL CYS GLN VAL THR HIS ARG PHE ARG ASP ASP
GAA GCA GCA TTC CAA GCT CAT TAC AAG GTA TGC CAA GTT ACT CAC AGA TTC AGG GAT GAT
1841
LEU LYS LYS THR VAL THR PRO GLN ASN PHE THR PRO GLY CYS TYR ARG THR LEU ASN LEU
CTA AAG AAA ACT GTT ACT CCT CAA AAT TTT ACA CCA GGA TGT TAC CGG ACA CTA AAT TTA
1901
PHE ARG TYR LYS SER ARG CYS TYR ILE PHE THR MET TRP ILE PHE LEU LEU VAL LEU GLU
TTT AGA TAC AAA AGC AGG TGC TAC ATC TTT ACA ATG TGG ATA TTT CTT CTT GTC TTA GAA
1961
SER ILE LEU TRP ALA ALA SER ALA SER GLU THR PRO LEU THR PRO VAL TRP ASN ASP ASN
TCC ATA CTG TGG GCT GCA AGT GCA TCA GAG ACA CCA TTA ACT CCT GTC TGG AAT GAC AAT
G2
2021
ALA HIS GLY VAL GLY SER VAL PRO MET HIS THR ASP LEU GLU LEU ASP PHE SER LEU THR
GCC CAT GGG GTA GGT TCT GTT CCT ATG CAT ACA GAT TTA GAG CTT GAT TTC TCT TTA ACA
2081
SER SER SER LYS TYR THR TYR ARG ARG LYS LEU THR ASN PRO LEU GLU GLU ALA GLN SER
TCC AGT TCC AAG TAT ACA TAC CGT AGG AAG TTA ACA AAC CCA CTT GAG GAA GCA CAG TCC
2141
ILE ASP LEU HIS ILE GLU ILE GLU GLU GLN THR ILE GLY VAL ASP VAL HIS ALA LEU GLY
ATT GAC CTA CAT ATT GAA ATA GAA GAA CAG ACA ATT GGT GTT GAT GTG CAT GCT CTA GGA
2201
HIS TRP PHE ASP GLY ARG LEU ASN LEU LYS THR SER PHE HIS CYS TYR GLY ALA CYS THR
CAC TGG TTT GAT GGT CGT CTT AAC CTT AAA ACA TCC TTT CAC TGT TAT GGT GCT TGT ACA
2261
LYS TYR GLU TYR PRO TRP HIS THR ALA LYS CYS HIS TYR GLU ARG ASP TYR GLN TYR GLU
AAG TAT GAA TAC CCT TGG CAT ACT GCA AAG TGC CAT TAT GAA AGA GAT TAC CAA TAT GAG
2321
THR SER TRP GLY CYS ASN PRO SER ASP CYS PRO GLY VAL GLY THR GLY CYS THR ALA CYS
ACG AGC TGG GGT TGT AAT CCA TCA GAT TGT CCT GGG GTG GGC ACA GGC TGT ACA GCA TGT
2381
GLY LEU TYR LEU ASP GLN LEU LYS PRO VAL GLY SER ALA TYR LYS ILE ILE THR ILE ARG
GGT TTA TAC CTA GAT CAA CTG AAA CCA GTT GGT AGT GCT TAT AAA ATT ATC ACA ATA AGG
2441
TYR SER ARG ARG VAL CYS VAL GLN PHE GLY GLU GLU ASN LEU CYS LYS ILE ILE ASP MET
TAC AGC AGG AGA GTC TGT GTT CAG TTT GGG GAG GAA AAC CTT TGT AAG ATA ATA GAC ATG
2501
ASN ASP CYS PHE VAL SER ARG HIS VAL LYS VAL CYS ILE ILE GLY THR VAL SER LYS PHE
AAT GAT TGT TTT GTA TCT AGG CAT GTT AAG GTC TGC ATA ATT GGT ACA GTA TCT AAA TTC
2561
SER GLN GLY ASP THR LEU LEU PHE PHE GLY PRO LEU GLU GLY GLY GLY LEU ILE PHE LYS
TCT CAG GGT GAT ACC TTA TTG TTT TTT GGA CCG CTT GAA GGT GGT GGT CTA ATA TTT AAA

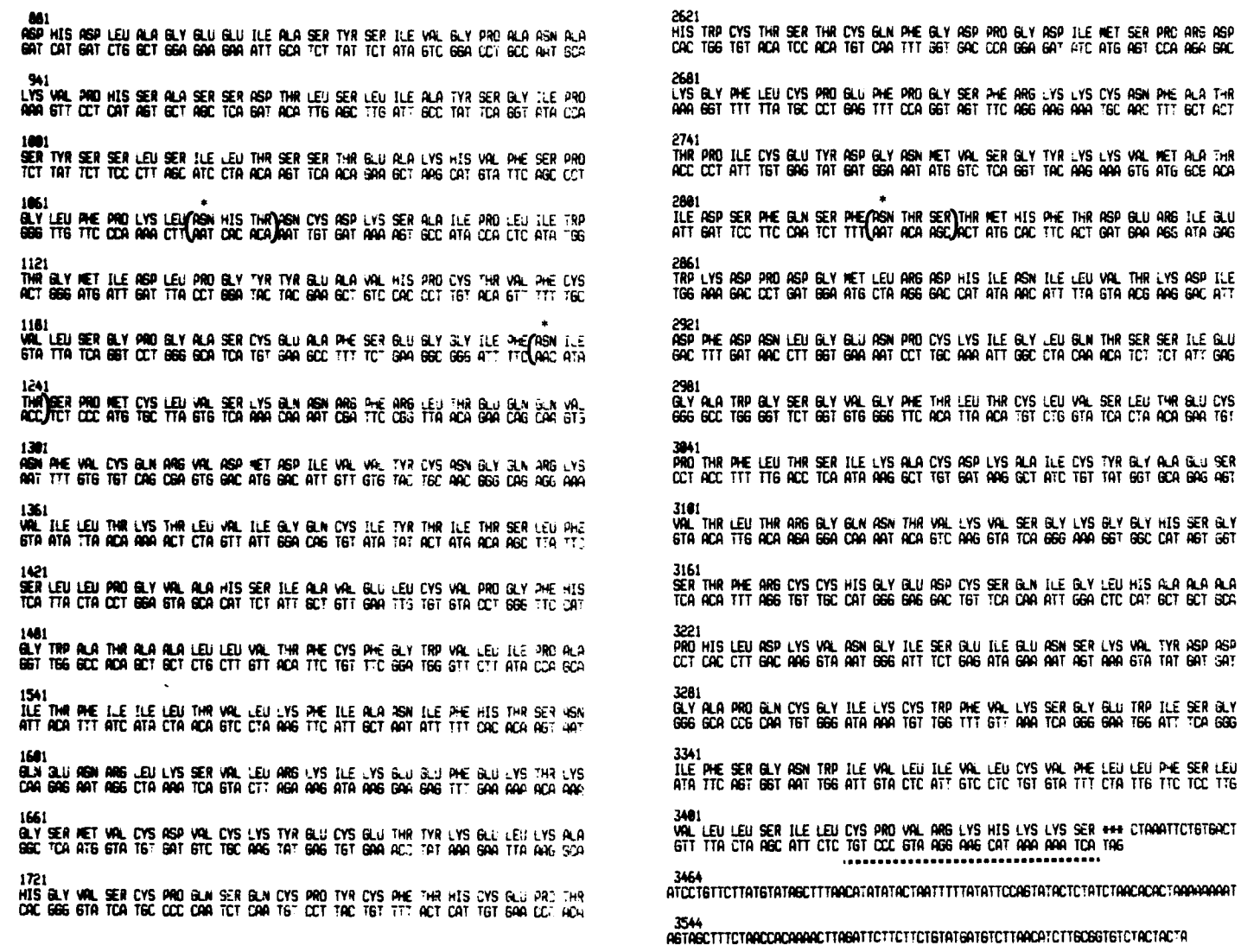


FIG. 2. Nucleotide sequence of the viral complementary-sense Hantaan M RNA segment and deduced amino acid sequence from the first encoded methionine of the open reading frame to the first stop codon. Amino acids identified by protein sequencing at the amino termini of G1 and G2 are underlined. Seven potential glycosylation sites (Asn X Ser/Thr) predicted from the nucleotide sequence are indicated by an *. Sequences corresponding to synthetic peptides used for definition of the carboxy termini of G1 and G2 are underscored with a dashed line.

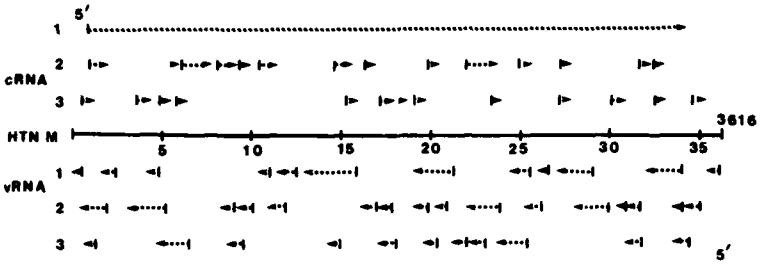


FIG. 3. Potential polypeptide coding regions of the six possible reading frames of viral (vRNA) or viral complementary-sense (cRNA) RNA. The solid line represents Hantaan M RNA in base pairs × 10^3.

plementary-sense RNA (Fig. 2). Examination of all six possible reading frames of the cDNA failed to reveal additional ORFs initiating at ATG codons and encoding 100 amino acids or more (Fig. 3). Within the major ORF, protein synthesis could potentially begin at either of two in-frame initiation codons at nucleotide positions 41–43 or 65–67. According to Kozak (1978, 1984), the first codon has more favorable flanking sequences for initiation of protein synthesis. A total of 1135 amino acids are encoded from the first ATG until a termination codon (TAG) occurs at nucleotide position 3346–3348. Similar to other Bunyaviridae, the putative gene product of the Hantaan M RNA was found to be relatively cysteine rich (Table 1) and of sufficient length to contain both the nonglycosylated G1 and G2 envelope glycoproteins, which have estimated M_r values of 62K and 52K, respectively (Table 1) (Collett *et al.*, 1985; Eshita *et al.*, 1985; Lees *et al.*, 1986; Schmaljohn *et al.*, 1986a).

In order to determine the gene order of the Hantaan M RNA segment and to locate the amino termini of the G1 and G2 glycoproteins, virion structural proteins were isolated and subjected to amino-terminal sequence analysis. A partial amino acid sequence of (NH_2)-Leu-X-X-Val-Tyr-Asp-Met-Lys-Ile-Glu-X-Pro-His-Thr-Thr-Val was determined for G1 and (NH_2)-X-Glu-X-Pro-Leu-X-Pro-Val-Trp-Asn-Asp-Asn-Ala-His-Gly-Val-Gly for G2, with X representing unidentified amino acids. These sequences correspond to those derived from cDNA beginning at nucleotide positions 95–97 (G1) and 1985–1987 (G2). Therefore, the gene order of the Hantaan M RNA, with respect to viral complementary-sense RNA, is 5′-G1-G2-3′.

A hydrophobicity/hydrophilicity plot of the potential gene product revealed that the amino terminus of G1 follows a hydrophobic stretch of 18 amino acids from the first ATG of the ORF (Fig. 4). Similarly, the amino terminus of G2, which originates at amino acid 649 of the ORF, follows a hydrophobic sequence. To estimate the position of the carboxy termini of G1 and G2, peptides corresponding to amino acids 588–614 (NH_2-Tyr-Lys-Val-Cys-Gln-Val-Thr-His-Arg-Phe-Arg-Asp-Asp-Leu-Lys-Lys-Thr-Val-Thr-Pro-Gln-Asn-Phe-Thr-Pro-Gly-Cys-COOH) and 1127–1135 (NH_2-Cys-Pro-Val-Arg-Lys-His-Lys-Lys-Ser-COOH) of the ORF were synthesized and antisera to each were generated in rabbits. The ability of the antipeptide sera to recognize G1 and G2 was measured by immune precipitation of native proteins and by hybridization to SDS-denatured proteins bound to nitrocellulose membranes. Immune precipitation of radiolabeled Hantaan virus

TABLE 1

AMINO ACID COMPOSITION OF HANTAAN M RNA GENE PRODUCTS[a]

Name	Number	Percentage
A—Alanine	52	4.6
C—Cysteine	61	5.4
D—Aspartic acid	50	4.4
E—Glutamic acid	58	5.1
F—Phenylalanine	58	5.1
G—Glycine	80	7.0
H—Histidine	37	3.3
I—Isoleucine	76	6.7
K—Lysine	67	5.9
L—Leucine	86	7.6
M—Methionine	23	2.0
N—Asparagine	40	3.5
P—Proline	51	4.5
Q—Glutamine	34	3.0
R—Arginine	34	3.0
S—Serine	98	8.6
T—Threonine	81	7.1
V—Valine	88	7.8
W—Tryptophan	19	1.7
Y—Tyrosine	42	3.7

[a] Deduced amino acids of the major open reading frame extending from nucleotides 41 to 3348. The predicted molecular weight of this polypeptide is 126,300.

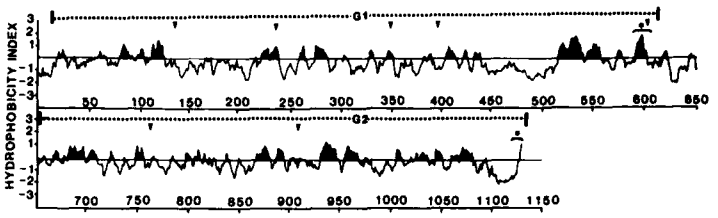


FIG. 4. A hydrophobicity/hydrophilicity plot of the gene product predicted from the cDNA sequence of viral complementary-sense Hantaan M RNA. Data points represent a running average taken over seven amino acid residues. Sequences with net hydrophilicity appear above the line and are shaded; sequences with net hydrophobicity appear below the line. Values assigned to amino acids are Ala, −0.5; His, −0.5; Arg, 3.0; Asp, 3.0; Cys, −1.0; Met, −1.3; Glu, 3.0; Lys, 3.0; Val, −1.5; Ile, −1.8; Ser, 0.3; Asn, 0.2; Leu, −1.8; Tyr, −2.3; Gln, 0.2; Gly, 0.0; Phe, −2.5; Trp, −3.4; Pro, 0.0; Thr, −0.4. The coding regions of G1 and G2 are overscored with a dashed line and the regions corresponding to synthetic peptides are marked with a *.

polypeptides from infected cell lysates demonstrated that antisera to peptide 588–614 specifically precipitated G1, and antisera to peptide 1127–1135 precipitated G2 (Fig. 5A). Specificities of the antipeptide sera were confirmed by probing blots of virion proteins electrophoretically transferred from polyacrylamide gels to nitrocellulose membranes. Rabbit antipeptide 1127–1135 sera reacted with G2 and antipeptide 588–614 sera reacted with G1 (Figs. 5B and 5C). Anti-Hantaan, polyclonal rabbit sera, and hyperimmune-mouse ascitic fluid were reactive with nucleocapsid protein, and to a lesser extent, G2, but not with G1. Preimmune rabbit sera, normal mouse ascitic fluid, and a mixture of monoclonal antibody-containing mouse ascitic fluids, directed against G1 or G2, did not react with any polypeptides on the immunoblots (Fig. 5C). Since the G2 peptide represents the exact 3′ nucleotide sequences of the ORF, these data indicate the G2 extends from amino acid 649 to 1127–1135 and has a calculated mol wt of 53,743. Although it was not possible to precisely define the carboxy terminus of G1, the immune-precipitation data indicate that G1 extends at least from amino acid 19 to 588–614 and, consequently, has an approximate mol wt of 64,000. Both of these values are consistent with estimated molecular weights for nonglycosylated Hantaan G1 and G2 previously reported (Schmaljohn *et al.*, 1986a).

Five potential asparagine-linked glycosylation sites (Asn–X–Ser/Thr) were identified in G1 and two in G2 (Figs. 2 and 3). These data are consistent with the more highly glycosylated nature of G1; however, we have not yet determined how many of these sites are actually used (Schmaljohn *et al.*, 1986a).

DISCUSSION

Nucleotide sequence analysis of the M genome segment of Hantaan virus revealed that Hantaan shares the common Bunyaviridae property of encoding a single gene product in the viral complementary-sense RNA. Although we have not demonstrated the occurrence of a polyprotein precursor to the glycoproteins in infected cells, we assume that both envelope glycoproteins are processed from this gene product, either by cotranslational or post-translational cleavage. The gene order of Hantaan M RNA with respect to messenger-sense RNA is 5′-G1-G2-3′. Hantaan apparently shares the common (Bunyaviridae) property of a leader sequence before the first encoded glycoprotein, as has been observed for the phlebo- and bunyaviruses (Collett *et al.*, 1985; Ihara *et al.*, 1985; Eshita *et al.*, 1984; Lees *et al.*, 1986). The 18 amino acids following the first AUG of the Hantaan M segment ORF and preceding the amino terminus of mature G1 constitute a typical glycoprotein signal sequence (von Heijne, 1983). The absence of a "NS_M" coding region between the signal sequence and G1 distinguishes Hantaan from the phleboviruses Punta Toro and Rift Valley fever viruses, which have long (30K and 16K, respectively) stretches of polypeptides prior to the amino-terminal sequences of their first mature glycoprotein (Collett *et al.*, 1985; Ihara *et al.*, 1985).

Sequence information for a polypeptide of approximately 70.3 k Da extends from the region between the amino terminus of G1 to the amino terminus of G2 and is considerably more than the 61K previously reported for nonglycosylated G1 (Schmaljohn *et al.*, 1986a). This discrepancy suggests that a small intergenic region between the carboxy terminus of G1 and the amino terminus of G2, similar to that predicted for Rift Vally fever virus, may be present (Collett *et al.*, 1985). Data presented in this paper demonstrate that the carboxy terminus of G1 extends at least to amino acid residues 588–614 and thus limits the size of the proposed intergenic region to less than 6K. Within this putative

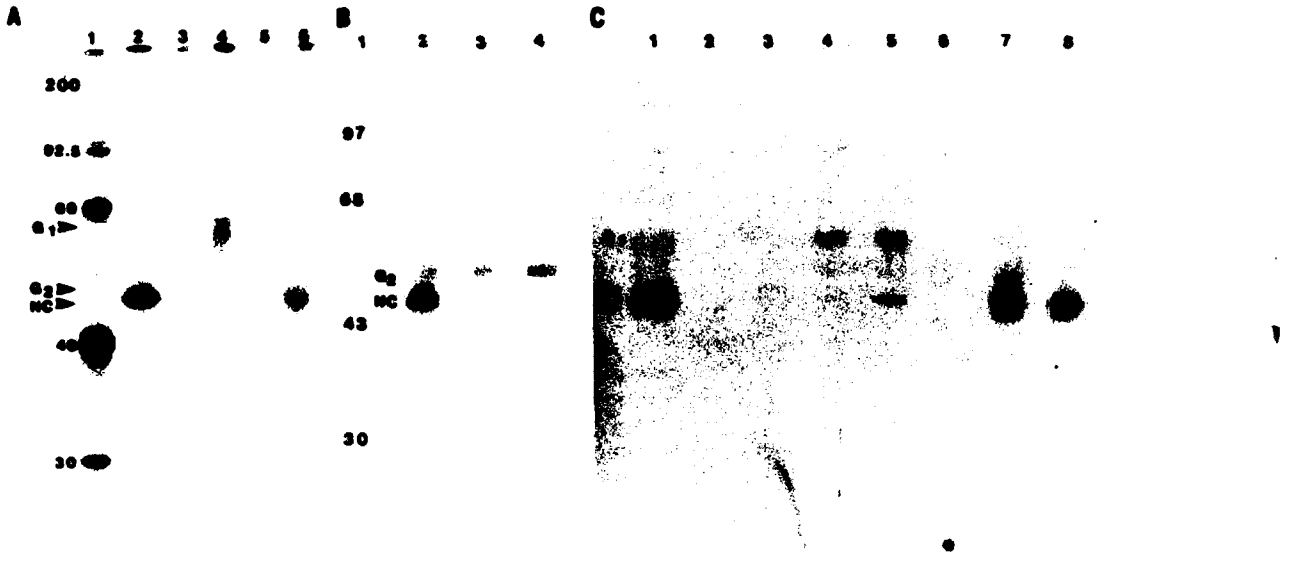


FIG. 5. (A) Polyacrylamide gel electrophoresis of polypeptides immune precipitated from Hantaan-infected cell lysates with antipeptide sera. Lane 1 contains ^{14}C-methylated protein markers of molecular weights indicated. Lane 2 contains Hantaan polypeptides precipitated with polyclonal anti-Hantaan sera. The more highly radiolabeled nucleocapsid protein (NC) migrates slightly faster than the G2 envelope glycoprotein. Lanes 3–6 contain Hantaan polypeptides immune precipitated with the following antipeptide immune sera: (3) preimmune peptide 588–614, (4) antipeptide 588–614, (5) preimmune peptide 1127–1135, (6) antipeptide 1127–1135. (B) Hantaan virion proteins electrophoretically transferred to nitrocellulose membranes and probed with immune sera. Numbers to the left of lane 1 correspond to migration of prestained molecular weight markers (Bethesda Research Laboratories) and are not visible by autoradiography. Lanes 2–4 were reacted with (2) polyclonal, anti-Hantaan rabbit sera, (3) antipeptide 1127–1135 (rabbit A), (4) antipeptide 1127–1135 (rabbit B). (C) Lane 1 contains Hantaan viral proteins radiolabeled with [^{35}S]methionine. Lanes 2–8 are immunoblots of virion proteins reacted with (2) normal mouse serum; (3) preimmune peptide 588–614; (4) antipeptide 588–614 (rabbit A); (5) antipeptide 588–614 (rabbit B); (6) mixture of five monoclonal antibody mouse ascitic fluids specific for Hantaan G1 or G2; (7) polyclonal anti-Hantaan rabbit sera; (8) hyperimmune anti-Hantaan mouse ascitic fluid.

intergenic region are amino acids which also conform to many of the parameters defining a signal sequence. In addition, amino acids 627–645 have a calculated average hydropathy value of 1.82 (Kyte and Doolittle, 1982), which is consistent with a membrane-spanning region. A similar 6K membrane-spanning intergenic region, which is thought to serve as a signal sequence, has been demonstrated between the two envelope protein coding sequences of the alphaviruses (Garoff *et al.*, 1980; Strauss and Strauss, 1986).

The carboxy terminus of G2 was readily defined, since synthetic peptides derived from the last coding sequences of the ORF were found to elicit anti-G2 antibodies in rabbits. The amino acids adjacent to the carboxy terminus of G2 (1107–1127) were found to be extremely hydrophobic, with an average hydropathy charge of 3.32, and were followed by a stretch of hydrophilic residues, suggesting the presence of a membrane-anchoring protein region. The precise location of membrane-associated regions of G1 and G2 remains to be determined.

Bunyaviridae generally mature intracellularly in the vicinity of the Golgi and infected cells do not usually express viral antigens on their plasma membranes (for a review see Bishop, 1985). Therefore, neutralizing antibodies which react with G1 and/or G2 may play a predominant role in recovery and immunity. A clear understanding of the mechanism of viral neutralization becomes especially relevant for the Hantaviruses because these viruses are unique among the Bunyaviridae in their ability to cause natural persistent infections of their rodent hosts and to be exclusively transmitted by aerosols (rather than by any known arthropod vector). For these reasons, it is important to determine the role of the envelope proteins in Hantaan viral infection and immunity. Presently, little is known about the function of G1 and G2, except that a monoclonal antibody to G1 can neutralize viral infectivity (unpublished data). We have examined the molecular properties of the genes encoding G1 and G2 and are continuing efforts to identify the functional characteristics of these proteins. These data should provide a basis for studies defining the molecular and antigenic nature of Hantaan viral infection and pathogenesis.

ACKNOWLEDGMENTS

We thank Dr. D. H. L. Bishop, Oxford University, for providing the synthetic oligonucleotide we used to prime first-strand synthesis; Dr. Francine McCucheon and David Roeder, MeLoy Laboratories, for nucleotide sequence determinations; William S. Lane, Department of Microsequencing, Harvard University, for amino-terminal sequence analysis of proteins; and Lynn Rasmussen for technical assistance.

REFERENCES

BISHOP, D. H. L. (1985). Replication of arenaviruses and bunyaviruses. *In* "Virology" (B. Fields, D. M. Knipe, R. M. Chanock, J. L. Melnick, B. Roizman, and R. E. Shope, Eds.), pp. 1083–1118. Raven Press, New York.

BISHOP, D. H. L., GOULD, K. G., AKASHI, H., and CLERX-VAN HAASTER, C. M. (1982). The complete sequence and coding content of snowshoe hare bunyavirus small (S) viral RNA species. *Nucleic Acids Res.* **10,** 3703–3713.

BISHOP, D. H. L., and SHOPE, R. E. (1979). Bunyaviridae. *In* "Comprehensive Virology" (H. Fraenkel-Conrat and R. R. Wagner, Eds.), Vol. 14, pp. 1–175. Plenum, New York.

CABRIDILLA, C. D., JR., HOLLOWAY, B. D., and OBIJESKI, J. F. (1983). Molecular cloning and sequencing of the LaCrosse virus S RNA. *Virology* **128,** 463–468.

COLLETT, M. S., PURCHIO, A. F., KEEGAN, K., FRAZIER, S., HAYS, W., ANDERSON, D. K., PARKER, M. D., SCHMALJOHN, C., SCHMIDT, J., and DALRYMPLE, J. M. (1985). Complete nucleotide sequence of the M RNA segment of Rift Valley fever virus. *Virology* **144,** 228–245.

DIETZSHOLD, B., LAFON, M., HUNG, H. W., OTVOS, L., and WUNNER, W. (1986). Domain structure of the N and Ns proteins of rabies virus. *J. Virol.*, submitted.

ELLIOT, L. H., KILEY, M. P., and MCCORMICK, J. B. (1984). Hantaan virus: Identification of virion proteins. *J. Gen. Virol.* **65,** 1285–1293.

ESHITA, Y., and BISHOP, D. H. L. (1984). The complete sequence of the M RNA of snowshoe hare bunyavirus reveals the presence of internal hydrophobic domains in the viral glycoprotein. *Virology* **137,** 227–240.

FULLER, F., and BISHOP, D. H. L. (1982). Identification of virus-coded nonstructural polypeptides in bunyavirus-infected cells. *J. Virol.* **41,** 643–648.

GAROFF, H., FRISCHAUF, A. M., SIMONS, K., LEHRACH, H., and DELIUS, H. (1980). Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. *Nature (London)* **288,** 236–241.

HOPP, T. P., and WOODS, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **78,** 3824–3828.

IHARA, T., AKASHI, H., and BISHOP, D. H. L. (1984). Novel coding strategy (ambisense genomic RNA) revealed by sequence analyses of Punta Toro phlebovirus S RNA. *Virology* **136,** 293–306.

IHARA, T., SMITH, J., DALRYMPLE, J. M., and BISHOP, D. H. L. (1985). Complete sequences of the glycoproteins and M RNA of Punta Toro phlebovirus compared to those of Rift Valley fever virus. *Virology* **144,** 246–259.

JOHNSON, D. A., GAUTSCH, J. W., SPORTSMAN, J. R., and ELDER, J. H. (1984). Improved technique utilizing nonfat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. *Gene Anal. Tech.* **1,** 3.

KOZAK, M. (1978). How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15,** 1109–1123.

KOZAK, M. (1984). Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin *in vivo*. *Nature (London)* **308,** 241–246.

KUISMANEN, E. (1984). Post-translational processing of Uukuniemi virus glycoproteins G1 and G2. *J. Virol.* **51,** 806–812.

KYTE, J., and DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132.

LEES, J. F., PRINGLE, C. R., and ELLIOTT, R. M. (1986). Nucleotide sequence of the Bunyamwera virus M RNA segment: Conservation of structural features in the Bunyavirus glycoprotein gene product. *Virology* **148,** 1–14.

LIU, F-T., ZINNECKER, M., HAMAOKA, T., and KATZ, D. H. (1979). New procedures for preparation and isolation of conjugates of proteins and a synthetic copolymer of D-amino acids and immunochemical characterization of such conjugates. *Biochemistry* **18,** 690–697.

MESSING, J. (1983). New M13 vectors for cloning. *In* "Methods in Enzymology" (R. Wu, L. Grossman, and K. Moldave, Eds.), Vol. **101,** pp. 20–78. Academic Press, Orlando, FL.

PARKER, M. D., and HEWLETT, M. J. (1981). The 3′ terminal nucleotide sequences of Uukuniemi and Inkoo virus RNA genome segments. *In* "Replication of Negative Strand Viruses" (D. H. L. Bishop and R. W. Compans, Eds.), pp. 125–145. Elsevier, New York.

SANGER, F., NICKLEN, S., and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.

SCHMALJOHN, C. S., and DALRYMPLE, J. M. (1983). Analysis of Hantaan virus RNA: Evidence for a new genus of Bunyaviridae. *Virology* **131,** 482–491.

SCHMALJOHN, C. S., and DALRYMPLE, J. M. (1984). Biochemical characterization of Hantaan virus. *In* "Segmented Negative Strand Viruses" (R. W. Compans and D. H. L. Bishop, Eds.), pp. 117–124. Elsevier, New York.

SCHMALJOHN, C. S., HASTY, S. E., DALRYMPLE, J. M., LEDUC, J. W., LEE, H. W., VON BONSDORFF, C.-H., BRUMMER-KORVENKONTIO, M., VAHERI, A., TSAI, T. F., REGNERY, H. L., GOLDGABER, D., and LEE, P. W. (1985). Antigenic and genetic properties of viruses linked to hemorrhagic fever with renal syndrome. *Science* **227,** 1041–1044.

SCHMALJOHN, C. S., HASTY, S. E., HARRISON, S. A., and DALRYMPLE, J. M. (1983). Characterization of Hantaan virions, the prototype virus of hemorrhagic fever with renal syndrome. *J. Infect. Dis.* **148,** 1005–1011.

SCHMALJOHN, C. S., HASTY, S. E., RASMUSSEN, L., and DALRYMPLE, J. M. (1986a). Hantaan virus replication: Effects of monensin, tunicamycin and endoglycosidases on the structural glycoproteins. *J. Gen. Virol.* **67,** 707–717.

SCHMALJOHN, C. S., JENNINGS, G. B., and DALRYMPLE, J. M. (1986b). Identification of Hantaan virus messenger RNA species. *In* "Segmented Negative Strand Viruses" (B. Mahy and D. Kolakofsky, Eds.) pp. 116–121. Elsevier, Amsterdam.

SCHMALJOHN, C. S., JENNINGS, G. B., HAY, J., and DALRYMPLE, J. M. (1986c). Coding strategy of the S-genome segment of Hantaan virus. *Virology* **155,** 633–643.

STEWART, J. M., and YOUNG, J. D. (1984). "Solid Phase Peptide Synthesis." Pierce Chemical Co. Rockford, IL.

STRAUSS, E. G., and STRAUSS, J. H. (1986). Structure and Replication of the alphavirus genome. *In* "The Togaviridae and Flaviviridae" (S. Schlesinger and M. J. Schlesinger, Eds.), pp. 35–90. Plenum, New York.

TAM, J. P., and MERRIFIELD, R. B. (1985). Solid phase synthesis of gastrin I. *Int. J. Pep. Protein Res.* **26,** 262–273.

ULMANEN, I., SEPPALA, P., and PETTERSSON, R. F. (1981). In vitro translation of Uukuniemi virus-specific RNAs: Identification of a nonstructural protein and a precursor to the membrane glycoproteins. *J. Virol.* **37,** 72–79.

VON HEIJNE, G. (1983). Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* **133,** 17–21.